sensors

MDPI

Article

Human Multi-Activities Classification Using mmWave Radar: Feature Fusion in Time-Domain and PCANet

Yier Lin [1,2], Haobo Li [3] and Daniele Faccio [4,*]

[1] School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; yierlin@foxmail.com
[2] Beijing Vocational College of Transport, Beijing 102618, China
[3] Department of Biomedical Engineering, School of Science and Engineering, University of Dundee, Dundee DD1 4HN, UK; hli005@dundee.ac.uk
[4] Extreme Light Group, School of Physics & Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
* Correspondence: daniele.faccio@glasgow.ac.uk

Abstract: This study introduces an innovative approach by incorporating statistical offset features, range profiles, time–frequency analyses, and azimuth–range–time characteristics to effectively identify various human daily activities. Our technique utilizes nine feature vectors consisting of six statistical offset features and three principal component analysis network (PCANet) fusion attributes. These statistical offset features are derived from combined elevation and azimuth data, considering their spatial angle relationships. The fusion attributes are generated through concurrent 1D networks using CNN-BiLSTM. The process begins with the temporal fusion of 3D range–azimuth–time data, followed by PCANet integration. Subsequently, a conventional classification model is employed to categorize a range of actions. Our methodology was tested with 21,000 samples across fourteen categories of human daily activities, demonstrating the effectiveness of our proposed solution. The experimental outcomes highlight the superior robustness of our method, particularly when using the Margenau–Hill Spectrogram for time–frequency analysis. When employing a random forest classifier, our approach outperformed other classifiers in terms of classification efficacy, achieving an average sensitivity, precision, F1, specificity, and accuracy of 98.25%, 98.25%, 98.25%, 99.87%, and 99.75%, respectively.

Keywords: human activity recognition; CNN-BiLSTM; mmWave; feature fusion; point cloud

check for updates

Citation: Lin, Y.; Li, H.; Faccio, D. Human Multi-Activities Classification Using mmWave Radar: Feature Fusion in Time-Domain and PCANet. *Sensors* **2024**, *24*, 5450. https://doi.org/10.3390/s24165450

Academic Editor: Guangcai Sun

Received: 15 July 2024
Revised: 19 August 2024
Accepted: 20 August 2024
Published: 22 August 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The World Health Organization has noted a significant growth in the global population of individuals aged 60 and above, from 1 billion in 2019 to an anticipated 1.4 billion by 2030 and further to 2.1 billion by 2050 [1]. Specifically, in China, the elderly population is expected to reach 402 million, making up 28% of the country's total population by 2040, due to lower birth rates and increased longevity [2]. This demographic shift has led to a rise in fall-induced injuries among the elderly, which is a leading cause of discomfort, disability, loss of independence, and early mortality. Statistics indicate that 28–35% of individuals aged 65 and above fall each year, which is a figure that increases to 32–42% for those aged over 70 [3]. As such, the remote monitoring of senior activities is becoming increasingly important.

Techniques for recognizing human activities fall into three categories: those that rely on wearable devices, camera systems, and sensor technologies [4]. Unlike wearables [5], camera- [6] and sensor-based methods offer the advantage of being fixed in specific indoor locations, eliminating the need for constant personal wear and thus better serving the home-bound elderly through remote monitoring while also addressing privacy concerns. Sensor-based technologies, in particular, utilize radio frequency signals' phase and amplitude without generating clear images, significantly reducing privacy violations compared to traditional methods.

Over the past decades, research in human activity recognition has explored various sensors. K. Chetty demonstrated the potential of using passive bistatic radar with WiFi for covertly detecting movement behind walls, enhancing signal clarity with the CLEAN method [7]. V. Kilaru explored identifying stationary individuals through walls using a Gaussian mixture model [8]. In our previous research, we developed a software-defined Doppler radar sensor for activity classification and employed various time–frequency image features, including the Choi-Williams and Margenau-Hill Spectrograms [9]. We proposed iterative convolutional neural networks with random forests [10] to accurately categorize several activities and individuals using FMCW radar, achieving successful activity recognition in diverse settings. Unfortunately, using the entire autocorrelation signal as input for the iterative convolutional neural networks with random forests led to excessive computational workload and frequent out-of-memory errors.

Thanks to numerous scattering centers, often called point clouds, mmWave radars are liable to provide high resolution. Because of their low cost and ease of use, mmWave radars have been gaining popularity. The team of A. Sengupta detected and tracked real-time human skeletons using mmWave radar in [11], while the team of S. An proposed a millimeter-based assistive rehabilitation system and a 3D multi-model human pose estimation dataset in [12,13]. Unfortunately, training fine-grained, accurate activity classifiers is challenging, as low-cost mmWave radar systems produce sparse, non-uniform point clouds.

This article introduces a cutting-edge classification algorithm designed to address the shortcomings of sparse and non-uniform point clouds generated by economical mmWave radar systems, thereby enhancing their applicability in practical scenarios. Our innovative solution boosts performance through the integration of statistical offset measures, range profiles, time–frequency analyses, and azimuth–range–time evaluations. It features nine distinct feature vectors composed of six statistical offset measures and three principal component analysis network (PCANet) fusion features derived from range profiles, time–frequency analyses, and azimuth–range–time imagery. These statistical offset measures are derived from I/Q data, which are harmonized through an angle relationship incorporating elevation and azimuth details. The fusion process involves channeling elevation and azimuth PCANet inputs through simultaneous 1D CNN-BiLSTM networks. This method prioritizes temporal fusion in the CNN-BiLSTM architecture for 3D range–azimuth–time data, followed by PCANet integration, enhancing the quality of fused features across the board and thereby improving overall classification accuracy.

The key contributions of this research are highlighted as follows:

(1) The introduction of an original method for classifying fourteen types of human daily activities, leveraging statistical offset measures, range profiles, time–frequency analyses, and azimuth–range–time data.
(2) Pioneering the use of the CNN-BiLSTM framework for fusing 3D range–azimuth–time information.
(3) Recommendation of the Margenau–Hill Spectrogram (MHS) for optimal feature quality and minimal feature count, which has been validated by analysis of four time–frequency methods.

In this paper, scalars and vectors are denoted by lowercase letters x and bold lowercase letters $\mathbf{x}$, whereas $=$ denotes the equal operator.

The remainder of this paper is structured as follows. Section 2 introduces the related works. Section 3 describes our methodology details using radar formula, feature source, feature fusion structure, and the framework of our method. Section 4 describes the experimental environment and the recording process of the measurement data. Our algorithm performance outcomes are illustrated with numerical results from the combinations classification of human daily action, which are also in Section 4. Future works and conclusions are drawn in Section 5.

2. Related Work

Numerous studies have focused on human activity recognition using millimeter wave (mmWave) technology. For instance, A. Pearce et al. [14] provided an in-depth review of the literature on multi-object tracking and sensing using short-range mmWave radar, while J. Zhang et al. [15] summarized the work on mmWave-based human sensing, and H. D. Mafukidze et al. [16] explored advancements in mmWave radar applications, offering ten public datasets for benchmarking in target detection, recognition, and classification.

Research employing feature-level fusion technology for human activity recognition includes a study in [17], which calculated mean Doppler shifts alongside micro-Doppler, statistical, time, and frequency domain features to feed into a support vector machine, achieving between 98.31% and 98.54% accuracy in classifying activities such as falling, walking, standing, sitting, and picking. E. Cardillo et al. [18] used range-Doppler and micro-Doppler features to classify different users' gestures. Given the challenges posed by sparse and irregular point clouds from budget-friendly mmWave radar, data enhancement techniques have become crucial for achieving high accuracy. A method proposed in [19] created samples with varied angles, distances, and velocities of human motion specifically for mmWave-based activity recognition. Another study in [20] utilized radar image classification and data augmentation, along with principal component analysis, to distinguish six activities, achieving 95.30% accuracy with the convolutional neural network (CNN) algorithm. E. Kurtoğlu et al. [21] exploited an approach that utilized multiple radio frequency data domain representations, including range-Doppler, time–frequency, and range–angle, for the sequential classification of fifteen American Sign Language words and three gross motor activities, achieving a detection rate of 98.9% and a classification accuracy of 92%.

Beyond feature-level fusion, autoencoders are widely used in mmWave signal analysis for activity recognition. The mmFall system in [22], utilizing a hybrid variational RNN autoencoder, reported a 98% success rate in detecting falls from 50 instances with only two false alarms. R. Mehta et al. [23] conducted a comparative study on extracting features through a convolutional variational autoencoder, showing the highest classification accuracy of 81.23% for four activities with the Sup-EnLevel-LSTM method.

Point cloud neural network technologies also play a pivotal role in recognizing human activities through mmWave signals. A real-time system in [24], using the PointNet model, demonstrated 99.5% accuracy in identifying five activities. The m-Activity system in [25] filtered human movements from background noise before classification with a specially designed lightweight neural network, resulting in 93.25% offline and 91.52% real-time accuracy for five activities. G. Lee and J. Kim leveraged spatial–temporal domain information for activity recognition using graph neural networks and a pre-trained model on point cloud and Kinect data [26], with their MTGEA model [27] achieving 98.14% accuracy in classifying various activities with mmWave and skeleton data.

Moreover, long short-term memory (LSTM) and CNN technologies have been essential in processing point clouds. C. Kittiyanpunya and team achieved 99.50% accuracy in classifying six activities using LSTM networks with 1D point clouds and Doppler velocity as inputs [28]. An end-to-end learning approach in [29] transformed each point cloud frame into 3D data for a custom CNN model, achieving a recall rate of 0.881. S. Khunteta et al. [30] showcased a technique where CNN extracted features from radio frequency images, followed by LSTM analyzing these features over time, reaching a peak accuracy of 94.7% for eleven activities. RadHAR in [31] utilized a time sliding window to process sparse point clouds into a voxelized format for CNN-BiLSTM classification, achieving 90.47% accuracy. Lastly, DVCNN in [32] improved data richness through radar rotation symmetry, employing a dual-view CNN to recognize activities, attaining accuracies of 97.61% and 98% for fall detection and a range of activities, respectively.

3. Methodology

This section includes four subsections. The first one introduces the standard formula of mmWave radar [33] for context. The second one displays where the information comes

from. The third one introduces the neural network structure for obtaining the fusion features. The last one presents the framework of our approach.

3.1. Radar Formula

In this article, mmWave radar with chirps was applied in human activity recognition. The transmission frequency was linearly increased over time through the transmit antenna. The chirp can be a formulated as follows:

$$S(t) = e^{j2\pi t(f_c + \frac{Bt}{2T_c}t)}, \quad 0 \le t \le T_c \tag{1}$$

where B, f_c, and T_c denote the bandwidth, carrier frequency, and chirp duration.

The time delay of the received signal has the following formula:

$$\tau = 2(R + vt)/c, \tag{2}$$

where R, v, and c denote the target range, the target velocity, and the light speed.

A radar frame is a consecutive chirp sequence, which is likely to be reshaped to a two-dimensional waveform across fast and slow time dimensions. There are M chirps with a sampling period T_{rep} (slow time dimension) in a frame, and each chirp has N points (fast time dimension) with f_s as the sampling rate. Hence, the intermediate frequency across fast and slow time dimensions together can be approximated as

$$d(n, m) \approx e^{j2\pi[\frac{(f_r + f_d)n}{f_s} + f_d m T_{rep} + \frac{2f_c R}{c}]}, \tag{3}$$

where n and m denote the index of fast and slow time samples, respectively. The information can be extracted by fast Fourier transform (FFT) along the fast and slow temporal dimensions. The range frequency f_r and Doppler f_d frequencies can be expressed as

$$f_r = \frac{2BR}{cT_c}, \tag{4}$$

$$f_d = 2f_c v/c. \tag{5}$$

Thanks to the received signal of each antenna having a different phase, a linear antenna array can estimate the target's azimuth. The phase shift between received signals from two adjacent antennas can be expressed as

$$\Delta\phi = 2\pi \frac{d \sin\theta}{\lambda}, \tag{6}$$

where θ denotes the target azimuth, while d denotes the distance between adjacent antennas. $\lambda = c/f_c$ denotes the base wavelength of the transmitted chirp.

For I number of targets, the three-dimensional intermediate frequency signal can be approximated as

$$d(n, m, l) \approx \sum_{i=1}^{I} a_i e^{j2\pi[\frac{(frq + f_{dq})n}{f_s} + \frac{ld \sin\theta_q}{\lambda} + f_{dq} m T_{rep} + \frac{2f_c R_q}{c}]}, \tag{7}$$

where i indicates the receiving antenna's index, while a_i denotes the ith target's complex amplitude. The samples of the intermediate frequency signal can be arranged to form a 3D Raw Data Cube across slow time, fast time, and channel dimensions, wherein the FFT can be applied along for velocity, range, and angle estimation.

3.2. Feature Source

In this article, the features as the classifiers' inputs mainly come from four categories of sources, i.e., the offset parameters, range profiles, time–frequency, and range–azimuth–time.

3.2.1. Offset Parameters

Some physical features such as speed and variation rates [34] also are liable to be extracted by the traditional statistic methods using time domain or frequency domain data. This article calculated the offset parameters, including the mean, variance, standard

deviation, kurtosis, skewness, and central moment. The mean [35] measures the signal probability distribution central tendency. The variance [36] measures the distance between the signal and its mean. The standard deviation [37] measures the input signal's variation or dispersion. The kurtosis [38] measures the signal probability distribution tailedness. The skewness [39] measures the asymmetry of the signal probability distribution about its mean. The central moment [40] measures the moment of the signal probability distribution about its mean, and we applied a two-order central moment in the following computation. These six offsets have proven to be effective for classification and were also used in our previous research. Algorithm 1 presents the pseudocode used to compute offsets.

Algorithm 1 Offsets()

Input: Elevation, Azimuth
Output: Offset Parameters
$Signal = Elevation + Azimuth \cdot e^{-i\pi/2}$
$Mean = \text{mean}(Signal)$
$Std = \text{std}(Signal)$
$Skewness = \text{skewness}(Signal)$
$Kurtosis = \text{kurtosis}(Signal)$
$Var = \text{var}(Signal)$
$Moment = \text{moment}(Signal, 2)$

3.2.2. Range Profiles

The range profile represents the target's time domain response to a high-range resolution radar pulse. As shown in Figure 1, the range profiles of measured data without desampling, which will be introduced in Section 4.1, show two targets. In this figure, the upper line represents the falling subject, while the lower line represents the standing subject. This is because the mmWave radar provides vertical information, so the range of the falling subject is greater than that of the standing subject when both the base of the standing subject and the radar are at the same horizontal level.

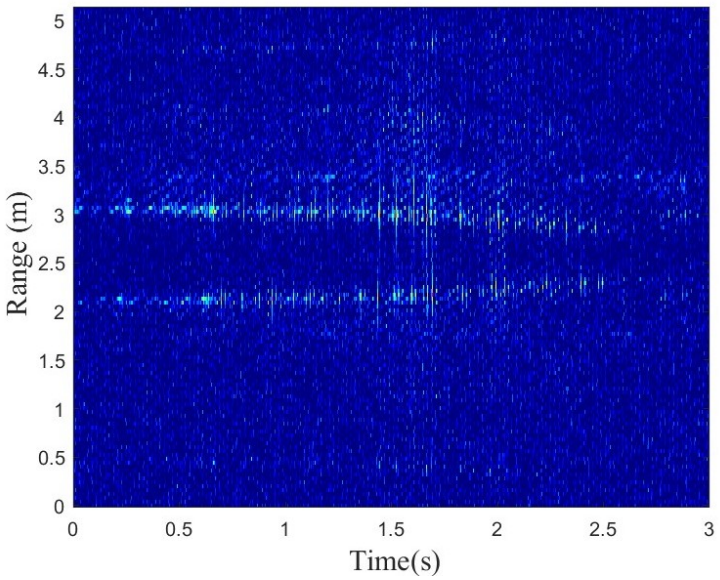


Figure 1. This figure shows the range profiles of measured data with two subjects doing standing (lower) and falling (upper).

3.2.3. Time-Frequency

Thanks to micro-Doppler signatures, different activities generate uniquely distinctive spectrogram features. Therefore, time-frequency analysis is significant for feature extraction.

The most common time-frequency analysis approach is the short-time Fourier transform (STFT) in Figure 2. Moreover, in addition to the STFT, three additional time–frequency distributions (Margenau-Hill Spectrogram [41], Choi-Williams [42], smoothed pseudo Wigner-Ville [43]) and their contributions in the final recognition will be compared in Section 4.

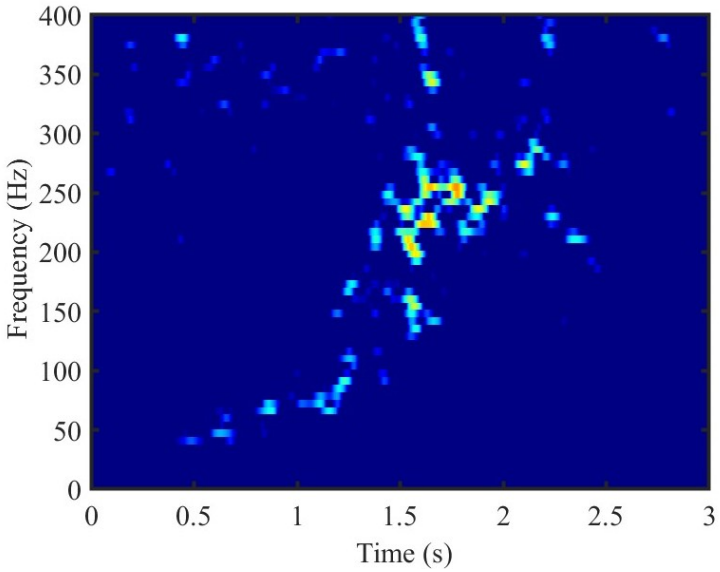


Figure 2. This figure displays the STFT image of measured data in Figure 1 with 800 Hz sampling frequency.

3.2.4. Range–Azimuth–Time

Range–azimuth–time plays a pivotal role in point cloud extraction, particularly in the context of object detection [44]. The range–azimuth dimensions, which are polar coordinates in Figure 3, are often converted into Cartesian coordinates to enhance their interpretability. The following Equations (8) formulated the coordinate transformation from the range and azimuth domain $[r, \theta]$ to the Cartesian domain represented by $[\mathbf{x}, \mathbf{y}]$.

$$\begin{cases} \mathbf{x} = \mathbf{r}\cos(\theta) \\ \mathbf{y} = \mathbf{r}\sin(\theta) \end{cases}, \quad \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]. \tag{8}$$

The instantaneous sparse point cloud can be drawn through the constant false alarm rate, as shown in Figure 4. If the time dimension is added to Figure 4, the mmWave radar data can be represented as 4D point clouds.

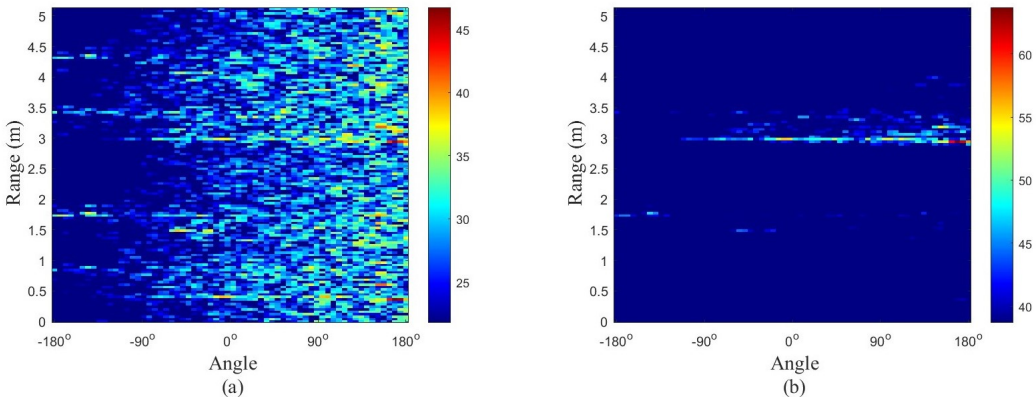


Figure 3. These images display the relationships between (**a**) horizontal and (**b**) vertical range and angles using elevation and azimuth information of measured data in Figure 1.

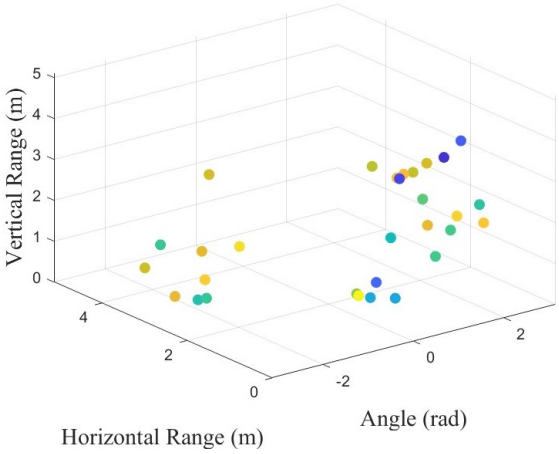


Figure 4. This figure shows the instantaneous 3D point cloud of measured data in Figure 3.

To reduce the number of features for the final recognition, the time-frequency and range-azimuth images should be analyzed and fused. Orthogonal and linear PCA transform input signals into a new coordinate system to extract main features and reduce the computational complexity [45,46]. The most important variance lies in the first coordinate, while the second most important variance lies in the second. Hence, the first PCA value is the most significant of the whole PCA value, and the following PCA value calculation is based on the former one. Moreover, CNN-BiLSTM has a structure that is BiLSTM. Thereby, the CNN-BiLSTM can fuse the PCANet. We applied PCA to analyze the images and CNN-BiLSTM to fuse the PCANet in this study.

3.3. CNN-BiLSTM

CNN-BiLSTM is a network that can reduce and eliminate the network for noise and dimensionality in data using a parallel structure for reducing the time complexity and an attention mechanism for promoting high accuracy via the key representations' weights redistribution [47,48].

As shown in Figure 5, the values of PCANet are fed into parallel 1D CNN networks, i.e., Stream A and Stream B. The parallel outputs of CNN streams will be multiplied based on the element. After unfolding and flattening, the multiplied sequence will be the input of a bidirectional LSTM (BiLSTM) structure. The BiLSTM structure includes two LSTM networks. The first LSTM network is for forward learning from the previous values, while the second LSTM network is for inverse learning from the upcoming values. The learned information will be combined in the attention layer. The attention layer has four hidden fully connected layers and a softmax, which is used to merge the upstream layer's output and filter the significant representations out for recognition purposes, i.e., the fusion feature of the inputs. The pseudocode for computing CNN-BiLSTM with K-fold cross-validation is shown in Algorithm 2, and its options are listed in Table 1.The usage of trainNetwork is shown in Figure 6, and its analysis is listed in Table 2.

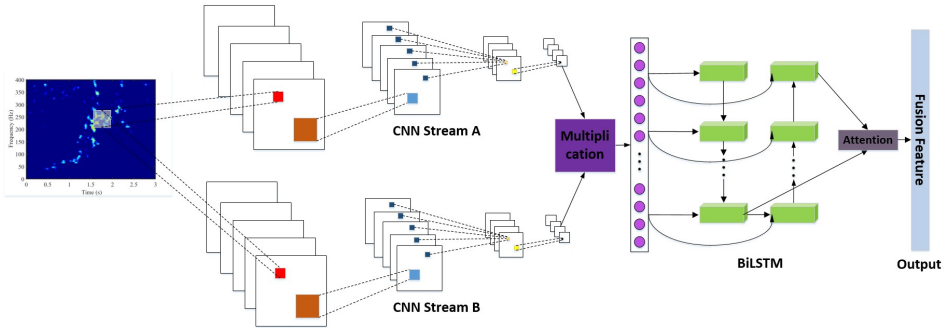


Figure 5. This figure shows the CNN-BiLSTM structure.

Algorithm 2 CNN-BiLSTM()

Input: FusionInput, KFoldNum
Output: Fusion Feature

Create CNN-BiLSTM Layers.
Set CNN-BiLSTM Options.
$COV = \text{cvpartition}(Label, 'KFold', KFoldNum);$
for $group = 1 : KFoldNum$ **do**
  $Index_Train(group, :) = \text{find}(\text{training}(COV, group));$
  $Index_Test(group, :) = \text{find}(\text{test}(COV, group));$
  $Train = InputFusion(Index_Train(group, :), :);$
  $Test = InputFusion(Index_Test(group, :), :);$
  $res = [Train; Test];$
  for $i = 1 : \text{length}(unique(Label))$ **do**
    $mid_res = res((res(:, \text{end}) == i), :);$
    $mid_size = \text{size}(mid_res, 1);$
    $mid_tiran = \text{round}(1 - 1/KFoldNum \times mid_size);$
    $P_train = [P_train; mid_res(1 : mid_tiran, 1 : \text{end} - 1)];$
    $T_train = [T_train; mid_res(1 : mid_tiran, \text{end})];$
    $P_test = [P_test; mid_res(mid_tiran + 1 : \text{end}, 1 : \text{end} - 1)];$
    $T_test = [T_test; mid_res(mid_tiran + 1 : \text{end}, \text{end})];$
  end for
  $net = \text{trainNetwork}(P_train, T_train, Layers, Options);$
  $T_sim1(group, :) = \text{vec2ind}(\text{predict}(net, P_train)');$
  $T_sim2(group, :) = \text{vec2ind}(\text{predict}(net, P_test)');$
end for
for $ii = 1 : group$ **do**
  $temp_T_sim1 = T_sim1(ii, :);$
  $temp_T_sim2 = T_sim2(ii, :);$
  for $jj = 1 : \text{size}(temp_T_sim1, 2)$ **do**
    $Temp(Index_Train(ii, jj)) = temp_T_sim1(jj);$
  end for
  for $jj = 1 : \text{size}(temp_T_sim2, 2)$ **do**
    $Temp(Index_Test(ii, jj)) = temp_T_sim2(jj);$
  end for
  $Fea(ii, :) = Temp;$
end for
$FuseFeature = \text{mean}(Fea, 1);$

Table 1. Option parameters of CNN-BiLSTM.

Optimizer	Parameters
Optimization	adam
Initial Learn Rate	0.001
ℓ_2 regularization	1×10^{-4}
Learn Rate Schedule	piecewise
Learn Rate Drop Factor	0.5
Learn Rate Drop Period	400
Shuffle	every epoch

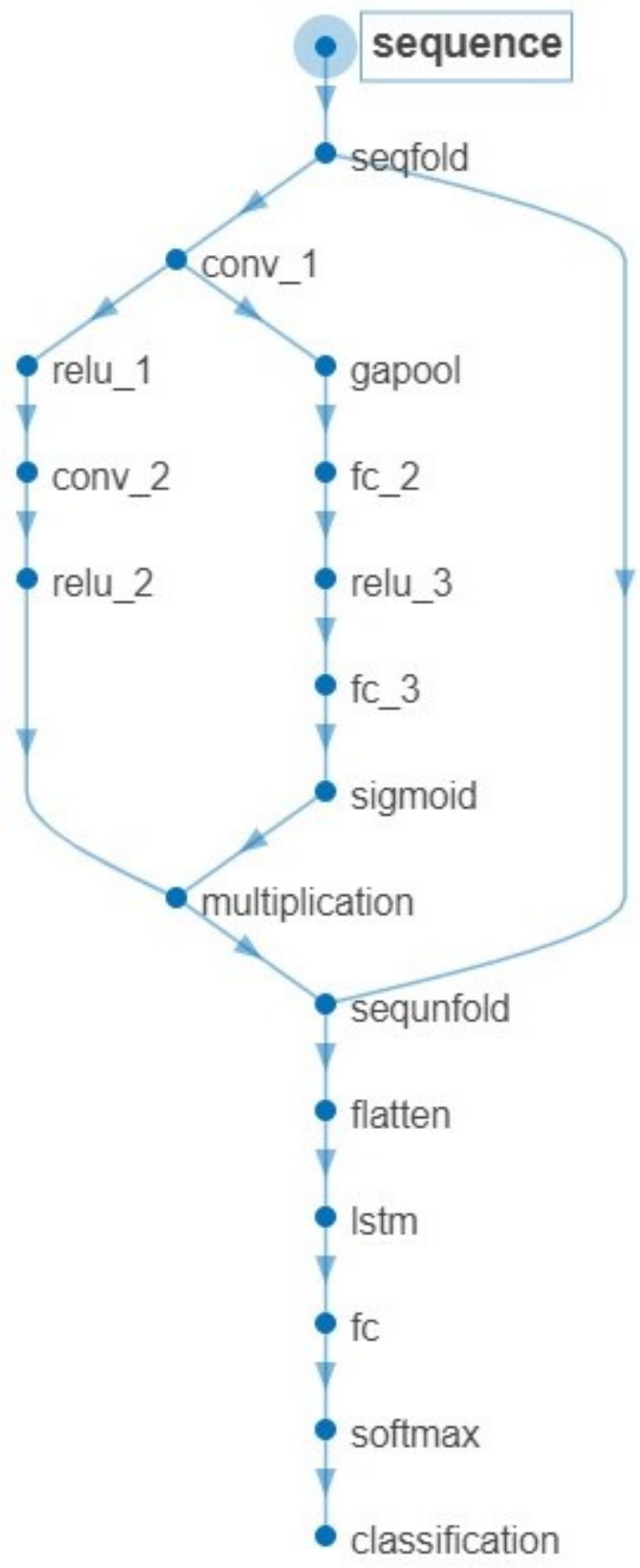


Figure 6. This figure displays the trainNetwork.

Table 2. Details of trainNetwork Usage in Figure 6.

Name	Type	Activations	Learnable Properties
sequence	Sequence Input	256(S) × 1(S) × 1(C) × 1(B) × 1(T)	-
seqfold	Sequence Folding	out 256(S) × 1(S) × 1(C) × 1(B) miniBatchSize 1(C) × 1(U)	-
conv_1	2D Convolution	254(S) × 1(S) × 32(C) × 1(B)	Weights 3 × 1 × 1 × 32 Bias 1 × 1 × 32
relu_1	ReLU	254(S) × 1(S) × 32(C) × 1(B)	-
conv_2	2D Convolution	252(S) × 1(S) × 64(C) × 1(B)	Weights 3 × 1 × 32 × 64 Bias 1 × 1 × 64
relu_2	ReLU	252(S) × 1(S) × 64(C) × 1(B)	-
gapool	2D Global Average Pooling	1(S) × 1(S) × 32(C) × 1(B)	-
fc_2	Fully Connected	1(S) × 1(S) × 16(C) × 1(B)	Weights 16 × 32 Bias 16 × 1
relu_3	ReLU	1(S) × 1(S) × 16(C) × 1(B)	-
fc_3	Fully Connected	1(S) × 1(S) × 64(C) × 1(B)	Weights 64 × 16 Bias 64 × 1
sigmoid	Sigmoid	1(S) × 1(S) × 64(C) × 1(B)	-
multiplication	Elementwise Multiplication	252(S) × 1(S) × 64 (C) × 1(B)	-
sequnfold	Sequence Unfolding	252(S) × 1(S) × 64(C) × 1(B) × 1(T)	-
flatten	Flatten	16,128(C) × 1(B) × 1(T)	-
lstm	BiLSTM	12(C) × 1(B)	InputWeights 48 × 16,128 RecurrentWeights 48 × 6 Bias 48 × 1
fc	Fully Connected	14(C) × 1(B)	Weights 14 × 12 Bias 14 × 1
softmax	Softmax	14(C) × 1(B)	-
classification	Classification Output	14(C) × 1(B)	-

3.4. Method Framework

Due to low-cost mmWave radar systems producing sparse, non-uniform point clouds—leading to low-quality feature extraction—training fine-grained, accurate activity classifiers is challenging. To solve this problem, fewer and more high-quality features for final classification are required to boost classification performance. Our novel approach for human activity classification, based on the statistical offset parameters, range profiles, time–frequency, and azimuth–range–time, is presented in Figure 7. Our method has nine fused feature vectors as the input of the classifier.

One of the key aspects of our method is the use of mmWave radar, which is capable of measuring elevation and azimuth information through its scanning method. We leveraged this capability by applying the angle relationship in space to fuse these two types of information, creating fused I/Q data. This data were then used for six statistical feature calculations: mean, variance, standard deviation, skewness, kurtosis, and two-order central moment.

Another one of the key aspects of our method is fusing PCANet images of range profiles and time–frequency. After the CNN-BiLSTM training, two fused vectors of range profiles and time–frequency can be obtained. In this part, we applied the elevation and azimuth data to obtain the PCANet instead of the angle relation fused I/Q data. The elevation and azimuth PCANet values were fed into parallel 1D networks of CNN-BiLSTM. We calculated the fusing PCANet features of images from range profiles or time–frequency using the pseudocode, as shown in Algorithms 3 and 4.

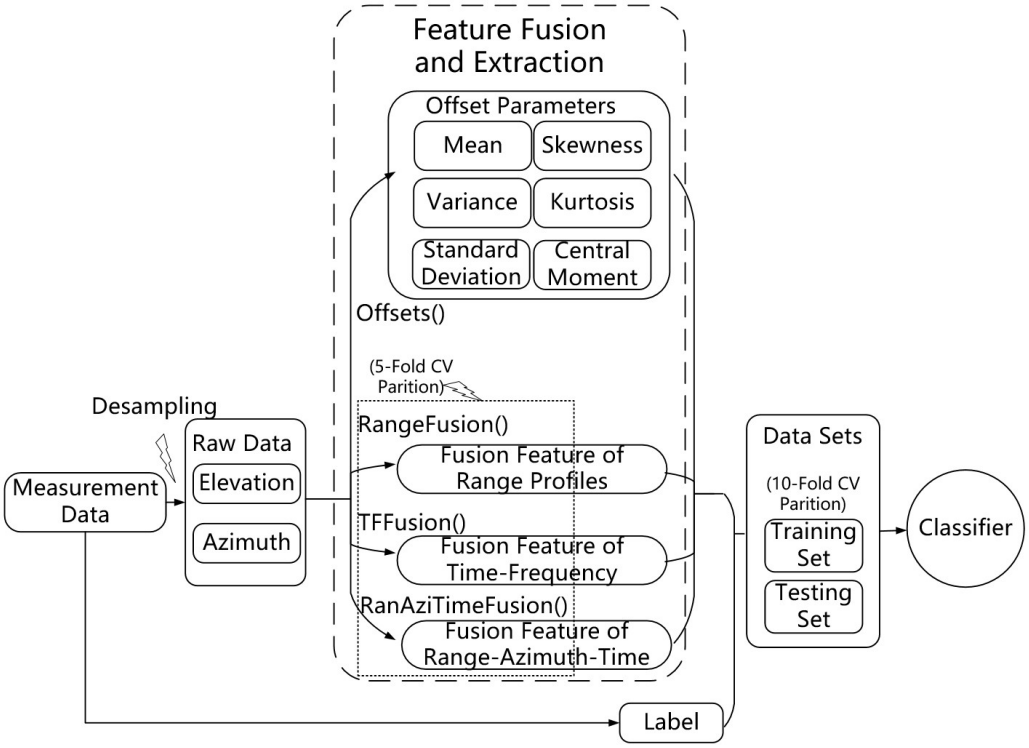


Figure 7. This figure shows the framework of our method.

Algorithm 3 RangeFusion()

Input: Elevation, Azimuth and Label of Sample Data, KFoldNum
Output: Fusion Feature of Range Profiles of Samples

Calculate Range Profiles of Elevation and Azimuth respectively.
Calculate PCANet of Range Profiles of Samples via SVD.
for $ii = 1 : \text{size}(PCA_Ele, 2)$ **do**
$FuseInput(:, (ii-1) \times 2 + 1) = PCA_Ele(:, ii)$;
$FuseInput(:, 2 \times ii) = PCA_Azi(:, ii)$;
$FuseInput(:, 2 \times \text{size}(PCA_Ele, 2) + 1) = Label$;
end for
$RangeFusion = \text{CNN-BiLSTM}(FuseInput, KFoldNum)$;

Algorithm 4 TFFusion()

Input: Elevation, Azimuth and Label of Sample Data, KFoldNum, Frebin, TFWin
Output: Fusion Feature of TF image of Samples

$Ele = \text{tfrmhs}(Elevation', 1 : \text{length}(Elevation), Frebin, TFWin)$;
$Azi = \text{tfrmhs}(Azimuth', 1 : \text{length}(Azimuth), Frebin, TFWin)$;
Calculate PCANet of Range Profiles of Samples via SVD.
for $ii = 1 : \text{size}(PCA_Ele, 2)$ **do**
$FuseInput(:, (ii-1) \times 2 + 1) = PCA_Ele(:, ii)$;
$FuseInput(:, 2 \times ii) = PCA_Azi(:, ii)$;
$FuseInput(:, 2 \times \text{size}(PCA_Ele, 2) + 1) = Label$;
end for
$TFFusion = \text{CNN-BiLSTM}(FuseInput, KFoldNum)$;

The third one of the key aspects of our method is fusing the PCANet of 3D range–azimuth–time. Because the actions are temporal processes, fusing 3D range–azimuth–time

temporally first can achieve higher quality features. The 3D range-azimuth-time fusion procedure was carried out on the CNN-BiLSTM structure with temporal fusion first, and PCANet fusion followed to ensure the best fusion, as shown in Algorithm 5. In this part, we calculated the range–azimuth–time using elevation and azimuth data instead of the angle relation fused I/Q data.

Algorithm 5 RanAziTimeFusion()

Input: Elevation, Azimuth and Label of Sample Data, KFoldNum
Output: Fusion Feature of Range–Azimuth–Time
Calculate Range–Azimuth–Time of Elevation and Azimuth, respectively.
Calculate the PCANet of Range–Azimuth–Time.
%PCAofRAT($SampleNum, FrameNum, PCANum$)
for $ii = 1$ to $FrameNum$ **do**
  $Fusion(:, (ii-1)*2+1, :) = PCAofRAT_Ele(:, ii, :)$
  $Fusion(:, 2*ii, :) = PCAofRAT_Azi(:, ii, :)$
end for
for $ii = 1$ to $PCANum$ **do**
  $Input(:, :) = Fusion(:, :, ii)$
  $Temp = [Input, Label]$
  $FuseTime =$ CNN-BiLSTM$(Temp, KFoldNum)$
end for
$FuseInput = [FuseTime, Label]$
$RATFusion =$ CNN-BiLSTM$(FuseInput, KFoldNum)$

4. Experimental Results and Analysis

This section will display the experiment setup and data collection in the first part and the implementation details and performance analysis in the following parts.

4.1. Experiment Setup and Data Collection

Classifying actions by individual subjects is straightforward, but identifying combinations of actions presents challenges. The experiments in this paper aimed to classify these action combinations. In this article, two evaluation modules named AWR2243 and DCA1000 EVM were applied for the experiment. The AWR2243 is the ideal mmWave radar sensor because of its ease of use, low cost, low power consumption, high-resolution sensing, precision, and compact size, whose parameters are listed in Table 3.

For the experimental setting, an activity room (6 m $\times$ 8 m) within the Doctoral Building of the University of Glasgow was chosen to support ample ground for our experiments. The distances between the testee (1.5 m), glass wall, and AWR2243 are shown in Figure 8. The radar resolution of the azimuth and elevation were 15° and 30°, respectively. In the experiment, the azimuth angle was mainly concerned with the moving testee. Therefore, the angular separation should range from 20° to 40°, which can resolve the two subjects spatially. The horizontal and vertical radar fields of view (FoVs) are both 60°, giving the radar FoV a concical shape. In our experiments, all testing subjects were within the radar's FoV. However, if additional targets are present outside the FoV, they may appear as 'ghost' targets in the radar spectrogram. The further these targets are from the radar's FoV, the lower their signal-to-noise ratio (SNR) will be.

Nine males and nine females participated in the experiment, with participants' identities made confidential for privacy. Detailed participant information is available in Table 4.

Table 3. Parameters of experimental device.

Parameters	Value
Radar Name	AWR2243
Frequency Range	76–81 GHz
Carrier Frequency	77 GHz
Number of Receivers	4
Number of Transmitters	3
Number of Samples per Chirp	256
Number of Chirps per Frame	128
Bandwidth	4 GHz
Range Resolution	4 cm
ADC Sampling Rate (max)	45 Msps
MIMO Modulation Scheme	TDM
Interface Type	MIPI-CSI2, SPI
Rating	Automotive
Operating Temperature Range	−40 to 140 °C
TI Functional Safety Category	Functional Safety Compliant
Power Supply Solution	LP87745-Q1
Evaluation Module	DCA1000

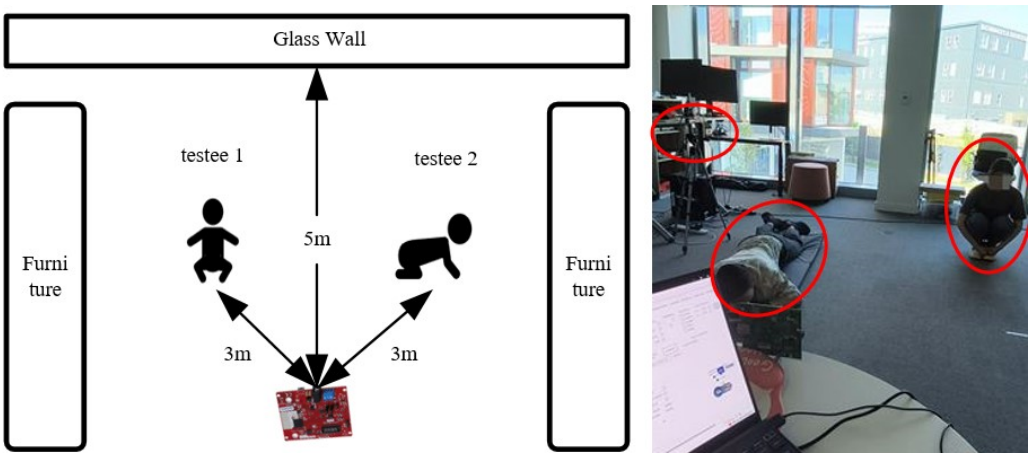


Figure 8. This figure displays the experimental scene diagram and photo.

Table 4. Information of the participants.

Participant ID	Gender	Height (cm)	Age
Participant 1	Male	172	25
Participant 2	Male	167	23
Participant 3	Male	165	24
Participant 4	Male	177	22
Participant 5	Male	174	26
Participant 6	Male	180	23
Participant 7	Male	185	25
Participant 8	Male	183	24
Participant 9	Male	188	22
Participant 10	Female	157	25
Participant 11	Female	155	23
Participant 12	Female	159	24
Participant 13	Female	167	22
Participant 14	Female	165	26
Participant 15	Female	170	23
Participant 16	Female	175	25
Participant 17	Female	173	24
Participant 18	Female	177	22

The experimental data were collected from fourteen categories of combinations of actions such as (I) bend and bend, (II) squat and bend, (III) stand and bend, (IV) walk and bend, (V) fall and bend, (VI) squat and squat, (VII) stand and squat, (VIII) fall and squat, (IX) walk and squat, (X) stand and stand, (XI) walk and stand, (XII) fall and stand, (XIII) walk and walk, and (XIV) fall and walk. There were 1500 samples in every category, with 21,000 samples in total. The time length of every sample was 3 s, with 800 Hz sampling after desampling processing. Hence, 21,000 was the total number of 3 s samples, while 2400 was the total number of frames of every sample, with the scatter plot shown in Figure 9. Moreover, the coherent processing interval (CPI) was 80 ms.

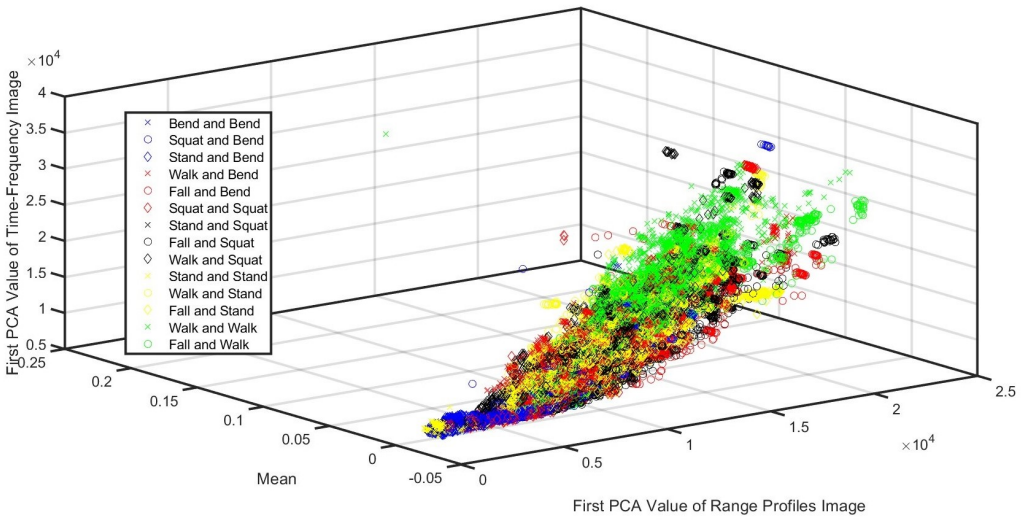


Figure 9. This figure shows the scatter plot of the fourteen categories of samples.

4.2. Implement Details

In this section, we applied five statistics (sensitivity, precision, F1, specificity, and accuracy) to measure the performance. These measures can be expressed as

$$Sensitivity = \frac{TP}{TP + FN}, \tag{9}$$

$$Precision = \frac{TP}{TP + FP}, \tag{10}$$

$$F1 = \frac{2TP}{2TP + FP + FN}, \tag{11}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{12}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{13}$$

where TP and FP are the number of true and false positives, while TN and FN are the number of true and false negatives. A true positive means a combination of actions is labeled correctly, a false positive means another combination of actions is labeled as the combination of actions, a true negative is a correct rejection, and a false negative is a missed detection.

Moreover, besides the STFT, three additional time–frequency distributions, named Margenau–Hill Spectrogram, Choi–Williams, and smoothed pseudo Wigner–Ville, contributed in the final recognition and will be compared in the following part. The four expressions of time–frequency can be written as

$$\boldsymbol{STFT_x(t, f; h)} = \iint_{-\infty}^{+\infty} \boldsymbol{x(u)H^*(u - t)}e^{(-j2\pi fu)}\boldsymbol{du}, \tag{14}$$

$$\begin{aligned} \boldsymbol{MHS_x(t,f)} &= \mathbb{R}\{\boldsymbol{K_{gh}^{-1}STFT_x(t,f;g)STFT_x^*(t,f;h)}\} \\ \boldsymbol{K_{gh}} &= \int \boldsymbol{h(u)g^*(u)du} \end{aligned}, \tag{15}$$

$$\boldsymbol{SPWV_x(t,\tau)} = \iint_{-\infty}^{+\infty} \boldsymbol{h(\tau)} \iint_{-\infty}^{+\infty} \boldsymbol{g(s-t)} \boldsymbol{x(s+\frac{\tau}{2})x^*(s-\frac{\tau}{2})ds}e^{-j2\pi v\tau}\boldsymbol{d\tau}, \tag{16}$$

$$\boldsymbol{CW_x(t,f)} = 2\iint_{-\infty}^{+\infty} \frac{\sqrt{\sigma}}{4\sqrt{\pi}|\boldsymbol{\tau}|}e^{-\frac{f^2\sigma}{16\tau^2}} \boldsymbol{x(t+f+\frac{\tau}{2})x^*(t+f-\frac{\tau}{2})}e^{-i2\pi\tau}\boldsymbol{dfdt}, \tag{17}$$

where $\boldsymbol{x}$ is the input signal, $\boldsymbol{t}$ represents the vectors of time instants, and $\boldsymbol{f}$ represents the normalized frequencies. σ denotes the square root of the variance. $\boldsymbol{STFT_x(\cdot)}$ is the short-time Fourier transform of $\boldsymbol{x}$. $\boldsymbol{h}$ is a smoothing window, while $\boldsymbol{g}$ is the analysis window.

In the feature extraction step, 5-fold cross-validation was employed to split the dataset into training and testing sets. Following the computation of five sets of fusion features, the results from these five groups were reconstructed into new datasets. Ultimately, the final classification outcomes for these new datasets were determined based on 10-fold cross-validation.

4.3. Performance Analysis

Figure 10 draws the classification sensitivity of every combination of actions of competitive temporal neural networks, such as the gated recurrent unit (GRU), LSTM, residual and LSTM recurrent networks (MSRLSTM), BiLSTM, attention-based BiLSTM (ABLSTM), temporal pyramid recurrent neural network (TPN), temporal CNN (TCN), and CNN-BiLSTM. The option parameters of these temporal neural network algorithms were all six hidden units, with Adam as the optimizer during training, a 0.0001 $\ell 2$ regularization, a 0.001 initial learn rate, a 0.5 drop factor learning rate, and 100 epochs. The time length of every sample was 3 s with 0.00125 s as the time interval. All the results are based on the average of 10-fold cross-validation. The average classification sensitivity for these action combinations is as follows: 34.56% for GRU, 35.57% for LSTM, 13.05% for MSRLSTM, 34.11% for BiLSTM, 34.95% for ABLSTM, 7.18% for TPN, 26.91% for TCN, and 33.69% for CNN-BiLSTM. All the algorithms in this figure serve as comparison methods for our proposed approach. Moveover, although CNN-BiLSTM did not have the highest average performance in temporal signal classification, we selected CNN-BiLSTM as the fusion method due to its data compatibility with the 3D fusion requirement in our algorithm.

As shown in Figure 7, the extracted features in our method came from the statistical offset parameters, range profiles, time–frequency, and azimuth–range–time. The offset parameters were based on the statistical calculation of the signal fused according to its elevation and azimuth information. The feature of range profiles was the output of the fusion of the PCANet of range profiles. The time–frequency feature was the output of the fusion of the PCANet of time–frequency images. The TF toolbox [49] analyzed the time–frequency, including the STFT, Margenau–Hill Spectrogram, Choi–Williams, and smoothed pseudo Wigner–Ville, i.e., (tfrstft.m, tfrmhs.m, tfrcw.m, and tfrspwv.m). The Frequency bin was 128, with a 127 hamming window. The feature of azimuth–range–time is the fusion output of both the temporal and PCANet of azimuth–range–time. Besides the offset parameters calculation, all the other feature fusion methods are based on CNN-BiLSTM. Moreover, all output results from CNN-BiLSTM in this paper are based on the 5-fold cross-validation to obtain the fusion output. Then, nine fused feature vectors were obtained as the classifier's inputs.

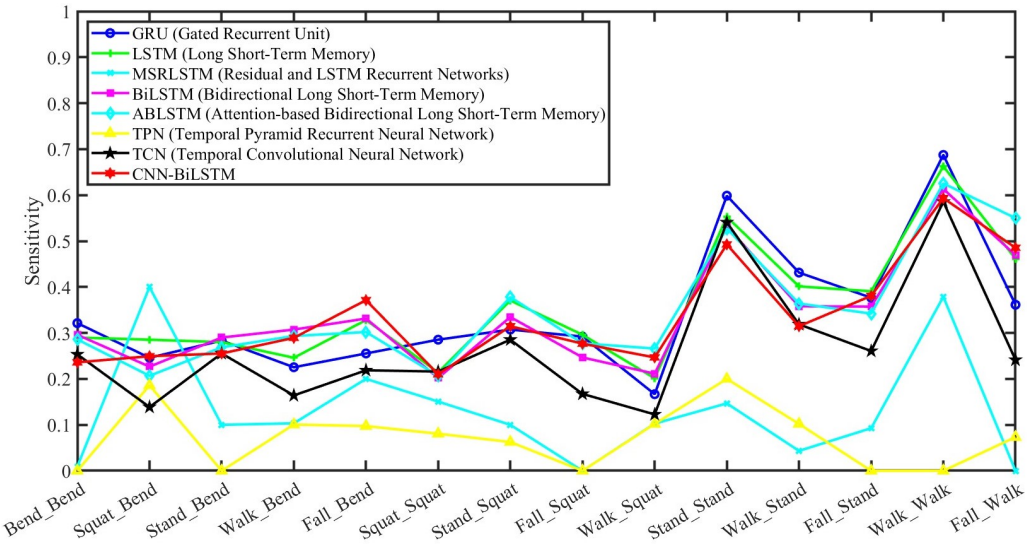


Figure 10. This figure shows the classification sensitivity of these combinations of actions using competitive temporal neural networks.

To find out the most suitable classifier for our approach, we utilized eight classifiers for the final classification, such as naïve Bayes (NB), pseudo quadratic discriminant analysis (PQDA), diagonal quadratic discriminant analysis (DQDA), K-nearest neighbors (K = 3), boosting, bagging, random forest (RF), and support vector machine (SVM). The SVM had the kernel of a two-order polynomial function with the auto-kernel scale, and the constraint was set to one with true standardization. The time length of every sample was 3 s, with 800 Hz sampling after desampling processing. Every group of 10-fold cross-validation was used by selecting 90% (18,900 samples) for learning features and the remaining 10% (2100 samples) for testing. All the results in this article are based on the average of all these 10 folds.

Figure 11 shows our average sensitivity of fourteen categories via various classifiers under the time–frequency condition of the STFT, Margenau–Hill Spectrogram, Choi–Williams, smoothed pseudo Wigner–Ville, and the joint of the above four methods, whose details are shown in Table 5. In CNN-BiLSTM processing, the epoch number was 100, and the fusion output was based on the average of five groups of 5-fold cross-validation. The final classification performance of all trials from the 10-fold partitions was the average of all ten groups. In this figure, the sensitivities from the bagging, random forest, and SVM surpassed 94%, while random forest, as the classifier, played better than the others. Thereby, random forest was subsequently considered as the unique classifier in the following research for convenience.

To evaluate the effect of the number of epochs of CNN-BiLSTM structure on the different time–frequency features for final classification performance, we tested the feature performance of our method under different epoch numbers with random forest as the classifier. The accuracies of these tests are displayed in Figure 12. In this figure, the joint of the four methods performed the best, followed by MHS. The vector number of the time–frequency feature was four, while that of the others was only one. The performance differences between the joint of four methods and MHS with 100 epochs came out to 0.15% sensitivity, 0.016% F1, 0.016% precision, 0.01% specificity, and 0.02% accuracy with 100 epochs. Hence, the time–frequency analysis method recommended in this article was MHS for fairness, which will be displayed in the following research of this paper for convenience.

Figure 13 displays the precision for classification using random forest as the classifier and MHS as the time–frequency analysis method. Details of this setup are listed in Table 6. The features in Test 1 and Test 2 include six offset parameters, 1 or 10 PCA values per range profile, and MHS images. In Test 3, the features consist of six offset parameters, the most significant PCA value of the range profile image, and the PCANet fusion of the MHS image. The key difference between Test 1 and Test 3 is whether the most significant PCA value or

the fused PCANet of the MHS image was used. In Test 4, the features include six offset parameters and PCANet fusions of both the range profile and MHS images. The difference between Test 3 and Test 4 is whether the feature set used the most significant PCA value or the PCANet fusion of the range profile images. Test 5 features include PCANet fusion features derived from range profiles, time–frequency analyses, and azimuth–range–time imagery. The difference between Test 4 and Test 6 (our method) is whether the feature vector came from the range–azimuth–time data. The difference between Test 5 and Test 6 is whether the feature vectors were derived from the offsets.

Table 5. Performance details of Figure 11.

Sensitivity	**STFT**	**MHS**	**CW**	**SPWV**	**Joint 4TFs**
NB	0.7380	0.7710	0.7449	0.7460	0.8066
PQDA	0.7926	0.8063	0.7889	0.7950	0.8376
DQDA	0.7414	0.7717	0.7458	0.7454	0.8062
KNN (k = 3)	0.5877	0.5875	0.5880	0.5861	0.5920
Boosting	0.4861	0.4863	0.4858	0.4862	0.4862
Bagging	0.9473	0.9595	0.9401	0.9424	0.9594
Random Forest	0.9771	0.9825	0.9754	0.9746	0.9840
SVM	0.9547	0.9651	0.9545	0.9550	0.9682
Precision	**STFT**	**MHS**	**CW**	**SPWV**	**Joint 4TFs**
NB	0.7663	0.7929	0.7691	0.7692	0.8176
PQDA	0.8127	0.8232	0.8105	0.8163	0.8503
DQDA	0.7691	0.7934	0.7700	0.7699	0.8170
KNN (k = 3)	0.5954	0.5952	0.5958	0.5938	0.5995
Boosting	0.3257	0.3399	0.3647	0.3970	0.3764
Bagging	0.9479	0.9598	0.9409	0.9429	0.9597
Random Forest	0.9772	0.9825	0.9755	0.9747	0.9841
SVM	0.9552	0.9655	0.9549	0.9554	0.9684
F1	**STFT**	**MHS**	**CW**	**SPWV**	**Joint 4TFs**
NB	0.7058	0.7453	0.7127	0.7140	0.7884
PQDA	0.7701	0.7879	0.7652	0.7712	0.8235
DQDA	0.7101	0.7459	0.7144	0.7139	0.7873
KNN (k = 3)	0.5859	0.5854	0.5861	0.5843	0.5900
Boosting	0.3570	0.3577	0.3748	0.3765	0.3681
Bagging	0.9471	0.9594	0.9399	0.9422	0.9594
Random Forest	0.9772	0.9825	0.9754	0.9746	0.9841
SVM	0.9548	0.9651	0.9546	0.9551	0.9682
Accuracy	**STFT**	**MHS**	**CW**	**SPWV**	**Joint 4TFs**
NB	0.9798	0.9824	0.9804	0.9805	0.9851
PQDA	0.9840	0.9851	0.9838	0.9842	0.9875
DQDA	0.9801	0.9824	0.9804	0.9804	0.9851
KNN (k = 3)	0.9683	0.9683	0.9683	0.9682	0.9686
Boosting	0.9605	0.9605	0.9604	0.9605	0.9605
Bagging	0.9959	0.9969	0.9954	0.9956	0.9969
Random Forest	0.9982	0.9987	0.9981	0.9980	0.9988
SVM	0.9965	0.9973	0.9965	0.9965	0.9976
Specificity	**STFT**	**MHS**	**CW**	**SPWV**	**Joint 4TFs**
NB	0.9626	0.9673	0.9636	0.9637	0.9724
PQDA	0.9704	0.9723	0.9698	0.9707	0.9768
DQDA	0.9631	0.9674	0.9637	0.9636	0.9723
KNN (k = 3)	0.9411	0.9411	0.9411	0.9409	0.9417
Boosting	0.9266	0.9266	0.9265	0.9266	0.9266
Bagging	0.9925	0.9942	0.9914	0.9918	0.9942
Random Forest	0.9967	0.9975	0.9965	0.9964	0.9977
SVM	0.9935	0.9950	0.9935	0.9936	0.9955

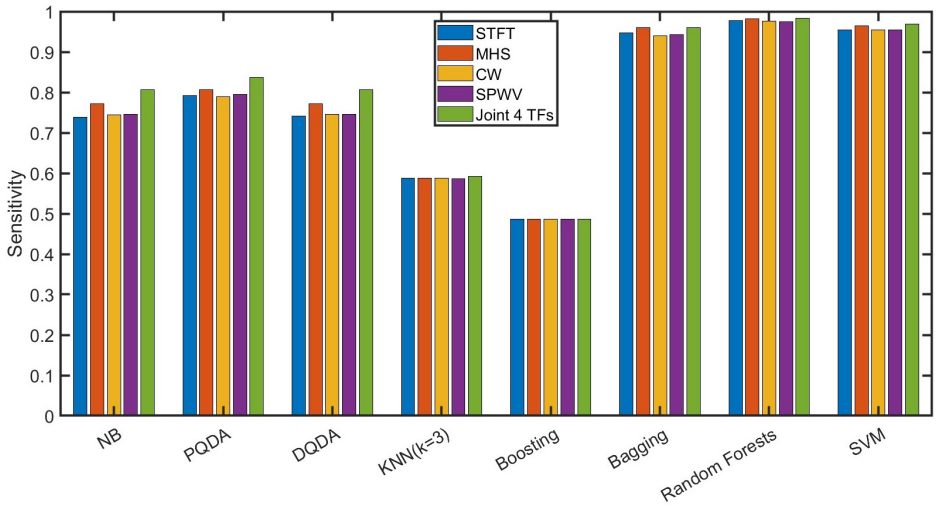


Figure 11. This figure shows our average sensitivity of fourteen categories via various classifiers under the time–frequency condition of STFT, Margenau–Hill Spectrogram, Choi–Williams, smoothed pseudo Wigner–Ville, and the joint of the above four methods.

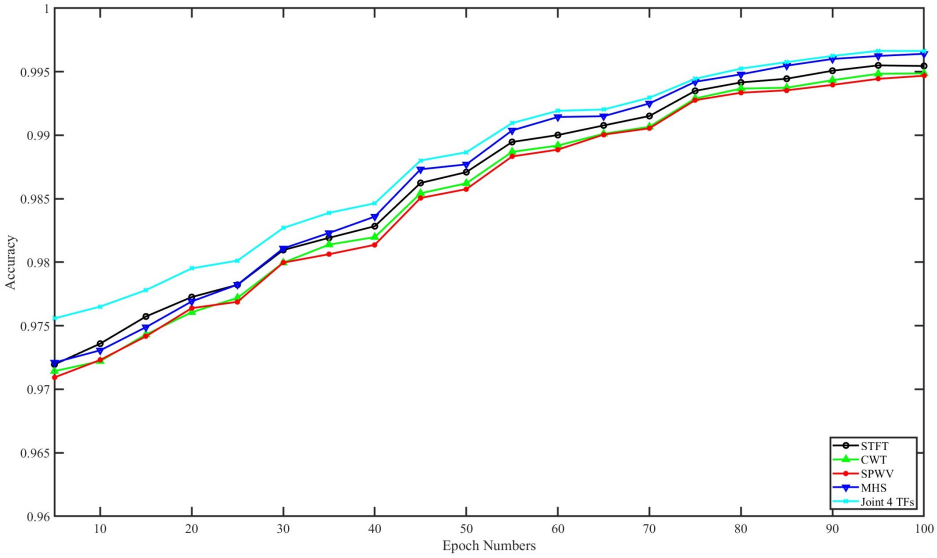


Figure 12. This figure shows the accuracy of our method under different epoch numbers.

In Table 6, it is shown that while the feature quantity could improve final classification performance, the feature quality had a greater impact on the boosting classification performance compared to the feature quantity. Table 7 presents the classification performance of random forest for each individual feature type from Table 6. Comparing the performance of Test 4 and Test 1, the features based on image PCANet fusion improved the final results by 10.92% in sensitivity, 11.01% in precision, 11.01% in F1, 0.84% in specificity, and 1.56% in accuracy. In Test 6, the offset feature vectors improved sensitivity, precision, F1, specificity, and accuracy by 1.34%, 1.34%, 1.34%, 0.10%, and 0.19%, respectively, compared to Test 5. Additionally, features derived from range–azimuth–time via temporal and PCANet fusion boosted sensitivity, precision, F1, specificity, and accuracy by 2.56%, 2.53%, 2.55%, 0.20%, and 0.37%, respectively, compared to Test 4. Tables 8 and 9 provide the confusion matrix and performance metrics for every action combination in Test 6.

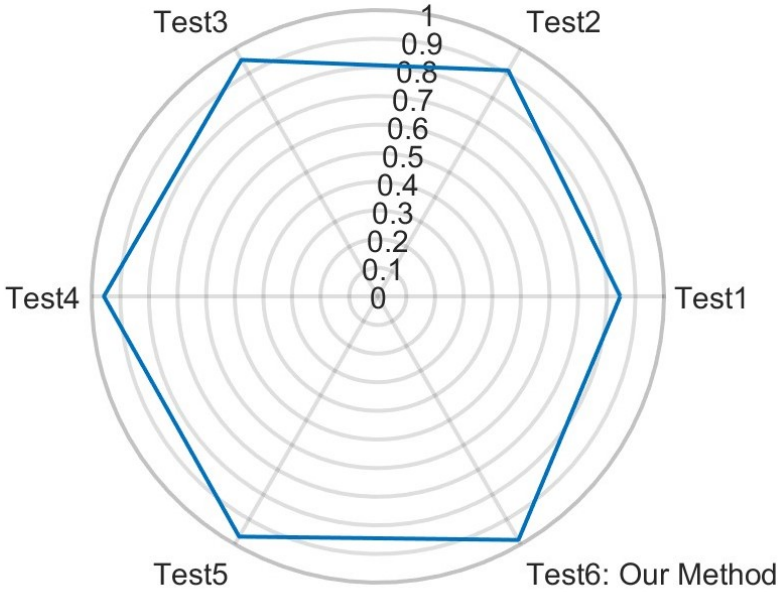


Figure 13. This figure shows the precision of our method (Test 6) and alternative method (Test 1–5).

Table 6. Performance fetails of Figure 13.

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Statistical Offsets	6	6	6	6	0	6
Range Features	1	10	1	1	1	1
TF Features	1	10	1	1	1	1
Range–Azimuth–Time	0	0	0	0	1	1
Total Features	8	26	8	8	3	9
Sensitivity	0.8477	0.9115	0.9535	0.9569	0.9691	0.9825
Precision	0.8472	0.9114	0.9538	0.9572	0.9692	0.9825
F1	0.8469	0.9112	0.9535	0.9570	0.9691	0.9825
Specificity	0.9883	0.9932	0.9964	0.9967	0.9976	0.9987
Accuracy	0.9782	0.9874	0.9934	0.9938	0.9956	0.9975

Table 7. Performance details of single-feature type of Table 6.

Single Feature Type	Feature Vectors	Sensitivity	Precision	F1	Specificity	Accuracy
Mean	1	0.0699	0.0701	0.0697	0.9285	0.8671
Variance	1	0.2605	0.2585	0.2588	0.9431	0.8944
Standard Deviation	1	0.2596	0.2576	0.2578	0.9430	0.8942
Kurtosis	1	0.2058	0.2055	0.2050	0.9389	0.8865
Skewness	1	0.1664	0.1670	0.1661	0.9359	0.8809
Central Moment	1	0.2485	0.2470	0.2471	0.9422	0.8926
Offset Parameters	6	0.2769	0.2763	0.2758	0.9444	0.8967

Table 7. *Cont.*

Single Feature Type	Feature Vectors	Sensitivity	Precision	F1	Specificity	Accuracy
PCA of Range Profiles	1	0.2769	0.2763	0.2758	0.9444	0.8967
PCA of Range Profiles	10	0.7675	0.7706	0.7663	0.9821	0.9668
PCANet Fusion of Range Profiles	1	0.6758	0.7048	0.6707	0.9751	0.9537
PCA of TF Image	1	0.3555	0.3551	0.3545	0.9504	0.9079
PCA of TF Image	10	0.7910	0.7960	0.7907	0.9839	0.9701
PCANet Fusion of TF Image	1	0.8937	0.8975	0.8941	0.9918	0.9848
Fusion of Range–Azimuth–Time	1	0.9253	0.9286	0.9251	0.9943	0.9893

Table 8. Confusion matrix of our method with STFT, 100 epochs, and random forest.

Action Combinations	ID	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Bend and Bend	**I**	**1465**	33	2	0	0	0	0	0	0	0	0	0	0	0
Squat and Bend	**II**	13	**1448**	18	1	1	11	4	0	0	4	0	0	0	0
Stand and Bend	**III**	1	21	**1468**	0	0	6	4	0	0	0	0	0	0	0
Walk and Bend	**IV**	0	0	0	**1475**	15	0	0	2	5	0	3	0	0	0
Fall and Bend	**V**	0	0	0	4	**1470**	19	0	6	0	0	1	0	0	0
Squat and Squat	**VI**	1	9	1	1	12	**1450**	15	7	0	4	0	0	0	0
Stand and Squat	**VII**	0	1	1	0	0	16	**1475**	4	0	3	0	0	0	0
Fall and Squat	**VIII**	0	1	0	0	6	6	17	**1455**	9	1	3	2	0	0
Walk and Squat	**IX**	0	0	0	4	1	0	0	20	**1469**	1	4	1	0	0
Stand and Stand	**X**	0	1	1	0	0	1	1	2	0	**1494**	0	0	0	0
Walk and Stand	**XI**	0	0	0	0	0	0	0	2	9	0	**1485**	4	0	0
Fall and Stand	**XII**	0	0	0	0	0	1	1	4	1	1	4	**1488**	0	0
Walk and Walk	**XIII**	0	0	0	0	0	0	0	0	0	0	0	3	**1497**	0
Fall and Walk	**XIV**	0	0	0	0	0	0	0	0	0	0	0	5	2	**1493**

Table 9. The performance of the confusion matrix in Table 8.

Action Combinations	Sensitivity	Precision	F1	Specificity	Accuracy
Bend and Bend	0.9767	0.9899	0.9832	0.9992	0.9976
Squat and Bend	0.9653	0.9564	0.9608	0.9966	0.9944
Stand and Bend	0.9787	0.9846	0.9816	0.9988	0.9974
Walk and Bend	0.9833	0.9933	0.9883	0.9995	0.9983
Fall and Bend	0.9800	0.9767	0.9784	0.9982	0.9969
Squat and Squat	0.9667	0.9603	0.9635	0.9969	0.9948
Stand and Squat	0.9833	0.9723	0.9778	0.9978	0.9968
Fall and Squat	0.9700	0.9687	0.9694	0.9976	0.9956
Walk and Squat	0.9793	0.9839	0.9816	0.9988	0.9974
Stand and Stand	0.9960	0.9907	0.9934	0.9993	0.9990
Walk and Stand	0.9900	0.9900	0.9900	0.9992	0.9986
Fall and Stand	0.9920	0.9900	0.9910	0.9992	0.9987
Walk and Walk	0.9980	0.9987	0.9983	0.9999	0.9998
Fall and Walk	0.9953	1.0000	0.9977	1.0000	0.9997
Average Performance	0.9825	0.9825	0.9825	0.9987	0.9975

5. Conclusions and Future Works

This paper introduces a pioneering method that utilizes statistical offset measures, range profiles, time–frequency analyses, and azimuth–range–time evaluations to effectively categorize various human daily activities. Our approach employs nine feature vectors, encompassing six statistical offset measures (mean, standard deviation, variance, skewness, kurtosis, and second-order central moments) and three principal component analysis network (PCANet) fusion attributes. These statistical measures were derived from combined elevation and azimuth data, considering their spatial angle connections. Fusion

processes for range profiles, time–frequency analyses, and 3D range–azimuth–time evaluations were executed through concurrent 1D CNN-BiLSTM networks, with a focus on temporal integration followed by PCANet application.

The effectiveness of this approach was demonstrated through rigorous testing, showcasing enhanced robustness particularly when employing the Margenau–Hill Spectrogram (MHS) for time–frequency analysis across fourteen distinct categories of human actions, including various combinations of bending, squatting, standing, walking, and falling. When tested with the random forest classifier, our method outperformed other classifiers in terms of overall efficacy, achieving impressive results: an average sensitivity, precision, F1, specificity, and accuracy of 98.25%, 98.25%, 98.25%, 99.87%, and 99.75%, respectively.

Research in radar-based human activity recognition has made strides, yet several promising areas remain in their infancy. Future directions include the following:

Radio-frequency-based activity recognition is known for minimizing privacy intrusions compared to traditional methods. Unlike camera-based systems that produce clear images through the combination of phase and amplitude in radio frequency signals, radio-frequency-based methods still raise privacy concerns, particularly when features could be linked to personal habits, or the handling of personal data requires careful consideration. An emerging field of study focuses on safeguarding user privacy while accurately detecting their activities.

For practical application, it is crucial to deploy activity recognition models in real time. This involves segmenting received signals into smaller sections for analysis. Addressing variables like window size and overlap during segmentation is vital. Moreover, training classification models presents unique challenges, especially in recognizing transitions between activities. Labeling these transitions and employing AI algorithms to fine-tune segmentation parameters represents a significant area for development.

Data augmentation has proven useful in image classification via generative adversarial networks (GANs). Applying GANs for augmenting time series data involves converting these data into images for augmentation and then back into time series format, addressing the issue of insufficient training data for classifiers—especially deep neural networks. Moreover, accurately labeling collected data without compromising privacy is challenging. While cameras are commonly used to verify data labels, they pose privacy risks, making unsupervised learning approaches that can cluster and label data without supervision increasingly relevant.

Moreover, the validation of our proposed method on other public radar datasets will be part of future work.

Author Contributions: Conceptualization, Y.L.; data curation, H.L.; formal analysis, Y.L.; funding acquisition, H.L. and D.F.; investigation, Y.L.; methodology, Y.L.; resources, H.L. and D.F.; software, Y.L.; validation, Y.L.; writing—original draft, Y.L. and H.L.; writing—review and editing, Y.L., H.L. and D.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the EPSRC IAA OSVMNC project (EP/X5257161/1).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABLSTM	Attention-based Bidirectional Long Short-term Memory
BiLSTM	Bidirectional Long Short-term Memory
CPI	Coherent Processing Interval
CNN	Convolutional Neural Network
CNN-BiLSTM	Convolutional Neural Network with Bidirectional Long Short-term Memory

CW	Choi–Williams
DQDA	Diagonal Quadratic Discriminant Analysis
FFT	Fast Fourier Transform
FoV	Radar Field of View
GANs	Generative Adversarial Networks
GRU	Gated Recurrent Unit
LSTM	Long Short-term Memory
MHS	Margenau–Hill Spectrogram
mmWave	Millimeter Wave
MSRLSTMs	Residual and Long Short-term Memory Recurrent Networks
NB	Naïve Bayes
PCA	Principal Component Analysis
PCANet	Principal Component Analysis Network
PQDA	Pseudo Quadratic Discriminant Analysis
RF	Random Forest
RNN	Recurrent Neural Network
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TCN	Temporal Convolutional Neural Network
TF	Time–Frequency
TPN	Temporal Pyramid Recurrent Neural Network

References

1. WHO. Ageing. Available online: https://www.who.int/health-topics/ageing (accessed on 15 July 2024).
2. WHO China. Ageing and Health in China. Available online: https://www.who.int/china/health-topics/ageing (accessed on 15 July 2024).
3. Ageing and Health (AAH); Maternal, Newborn, Child & Adolescent Health & Ageing (MCA). WHO Global Report on Falls Prevention in Older Age. WHO. Available online: https://www.who.int/publications/i/item/9789241563536 (accessed on 15 July 2024).
4. Wu, J.; Han, C.; Naim, D. A Voxelization Algorithm for Reconstructing mmWave Radar Point Cloud and an Application on Posture Classification for Low Energy Consumption Platform. *Sustainability* **2023**, *15*, 3342. [CrossRef]
5. Luwe, Y.J.; Lee, C.O.; Lim, K.M. Wearable Sensor-based Human Activity Recognition with Hybrid Deep Learning Model. *Informatics* **2022**, *9*, 56. [CrossRef]
6. Ullah, H.; Munir, A. Human Activity Recognition Using Cascaded Dual Attention CNN and Bi-Directional GRU Framework. *J. Imaging* **2023**, *9*, 130. [CrossRef]
7. Chetty, K.; Smith, G.E.; Woodbridge, K. Through-the-Wall Sensing of Personnel using Passive Bistatic WiFi Radar at Standoff Distances. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1218–1226. [CrossRef]
8. Kilaru, V.; Amin, M.; Ahmad, F.; Sévigny, P.; DiFilippo, D. Gaussian Mixture Modeling Approach for Stationary Human Identification in Through-the-wall Radar Imagery. *J. Electron. Imaging* **2015**, *24*, 013028. [CrossRef]
9. Lin, Y.; Le Kernec, J.; Yang, S.; Fioranelli, F.; Romain, O.; Zhao, Z. Human Activity Classification with Radar: Optimization and Noise Robustness with Iterative Convolutional Neural Networks Followed With Random Forests. *IEEE Sens. J.* **2018**, *18*, 9669–9681. [CrossRef]
10. Lin, Y.; Yang, F. IQ-Data-Based WiFi Signal Classification Algorithm Using the Choi-Williams and Margenau-Hill-Spectrogram Features: A Case in Human Activity Recognition. *Electronics* **2021**, *10*, 2368. [CrossRef]
11. Sengupta, A.; Jin, F.; Zhang, R.; Cao, S. mm-Pose: Real-Time Human Skeletal Posture Estimation using mmWave Radars and CNNs. *IEEE Sens. J.* **2020**, *20*, 10032–10044. [CrossRef]
12. An, S.; Ogras, U.Y. MARS: MmWave-based Assistive Rehabilitation System for Smart Healthcare. *ACM Trans. Embed. Comput. Syst.* **2021**, *20*, 1–22. [CrossRef]
13. An, S.; Li, Y.; Ogras, U.Y. mRI: Multi-modal 3D Human Pose Estimation Dataset using mmWave, RGB-D, and Inertial Sensors. *arXiv* **2022**, arXiv:2210.08394.
14. Pearce, A.; Zhang, J.A.; Xu, R.; Wu, K. Multi-Object Tracking with mmWave Radar: A Review. *Electronics* **2023**, *12*, 308. [CrossRef]
15. Zhang, J.; Xi, R.; He, Y.; Sun, Y.; Guo, X.; Wang, W.; Na, X.; Liu, Y.; Shi, Z.; Gu, T. A Survey of mmWave-based Human Sensing: Technology, Platforms and Applications. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 2052–2087. [CrossRef]
16. Mafukidze, H.D.; Mishra, A.K.; Pidanic, J.; Francois, S.W.P. Scattering Centers to Point Clouds: A Review of mmWave Radars for Non-Radar-Engineers. *IEEE Access* **2022**, *10*, 110992–111021. [CrossRef]
17. Muaaz, M.; Waqar, S.; Pätzold, M. Orientation-Independent Human Activity Recognition using Complementary Radio Frequency Sensing. *Sensors* **2023**, *13*, 5810. [CrossRef]

18. Cardillo, E.; Li, C.; Caddemi, A. Heating, Ventilation, and Air Conditioning Control by Range-Doppler and Micro-Doppler Radar Sensor. In Proceedings of the 18th European Radar Conference (EuRAD), London, UK, 5–7 April 2022; pp. 21–24. [CrossRef]
19. Wang, Z.; Jiang, D.; Sun, B.; Wang, Y. A Data Augmentation Method for Human Activity Recognition based on mmWave Radar Point Cloud. *IEEE Sens. Lett.* **2023**, *7*, 1–4. [CrossRef]
20. Taylor, W.; Dashtipour, K.; Shah, S.A.; Hussain, A.; Abbasi, Q.H.; Imran, M.A. Radar Sensing for Activity Classification in Elderly People Exploiting Micro-Doppler Signatures using Machine Learning. *Sensors* **2021**, *11*, 3881. [CrossRef]
21. Kurtoğlu, E.; Gurbuz, A.C.; Malaia, E.A.; Griffin, D.; Crawford, C.; Gurbuz, S.Z. ASL Trigger Recognition in Mixed Activity/Signing Sequences for RF Sensor-Based User Interfaces. *IEEE Trans. Hum. Mach. Syst.* **2022**, *52*, 699–712. [CrossRef]
22. Jin, F.; Sengupta, A.; Cao, S. mmFall: Fall Detection Using 4-D mmWave Radar and a Hybrid Variational RNN AutoEncoder. *IEEE Trans. Autom. Sci. Eng.* **2022**, *19*, 1245–1257. [CrossRef]
23. Mehta, R.; Sharifzadeh, S.; Palade, V.; Tan, B.; Daneshkhah, A.; Karayaneva, Y. Deep Learning Techniques for Radar-based Continuous Human Activity Recognition. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1493–1518. [CrossRef]
24. Alhazmi, A.K.; Mubarak, A.A.; Alshehry, H.A.; Alshahry, S.M.; Jaszek, J.; Djukic, C.; Brown, A.; Jackson, K.; Chodavarapu, V. P. Intelligent Millimeter-Wave System for Human Activity Monitoring for Telemedicine. *Sensors* **2024**, *24*, 268. [CrossRef]
25. Wang, Y.; Liu, H.; Cui, K.; Zhou, A.; Li, W.; Ma, H. m-Activity: Accurate and Real-Time Human Activity Recognition via Millimeter Wave Radar. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021), Toronto, ON, Canada, 6–11 June 2021; pp. 8298–8302. [CrossRef]
26. Lee, G.; Kim, J. Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-trained 3D Human-joint Coordinates. *Appl. Sci.* **2022**, *12*, 2168. [CrossRef]
27. Lee, G.; Kim, J. MTGEA: A Multimodal Two-Stream GNN Framework for Efficient Point Cloud and Skeleton Data Alignment. *Sensors* **2023**, *5*, 2787. [CrossRef]
28. Kittiyanpunya, C.; Chomdee, P.; Boonpoonga, A.; Torrungrueng, D. Millimeter-Wave Radar-Based Elderly Fall Detection Fed by One-Dimensional Point Cloud and Doppler. *IEEE Access* **2023**, *11*, 76269–76283. [CrossRef]
29. Rezaei, A.; Mascheroni, A.; Stevens, M.C.; Argha, R.; Papandrea, M.; Puiatti, A.; Lovell, N.H. Unobtrusive Human Fall Detection System Using mmWave Radar and Data Driven Methods. *IEEE Sens. J.* **2023**, *23*, 7968–7976. [CrossRef]
30. Khunteta, S.; Saikrishna, P.; Agrawal, A.; Kumar, A.; Chavva, A.K.R. RF-Sensing: A New Way to Observe Surroundings. *IEEE Access* **2022**, *10*, 129653–129665. [CrossRef]
31. Singh, A.D.; Sandha, S.S.; Garcia, L.; Srivastava, M. Human Activity Recognition from Point Clouds Generated through a Millimeter-wave Radar. In Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems, Los Cabos, Mexico, 25 October 2019; pp. 51–56. [CrossRef]
32. Yu, C.; Xu, Z.; Yan, K.; Chien, Y.R.; Fang, S.H.; Wu, H.C. Noninvasive Human Activity Recognition using Millimeter-Wave Radar. *IEEE Syst. J.* **2022**, *7*, 3036–3047. [CrossRef]
33. Ma, C.; Liu, Z. A Novel Spatial–temporal Network for Gait Recognition using Millimeter-wave Radar Point Cloud Videos. *Electronics* **2023**, *12*, 4785. [CrossRef]
34. Li, J.; Li, X.; Li, Y.; Zhang, Y.; Yang, X.; Xu, P. A New Method of Tractor Engine State Identification based on Vibration Characteristics. *Processes* **2023**, *11*, 303. [CrossRef]
35. Hippel, P.T.V. Mean, Median, and Skew: Correcting a Textbook Rule. *J. Stat. Educ.* **2005**, *13*. [CrossRef]
36. Asfour, M.; Menon, C.; Jiang, X. A Machine Learning Processing Pipeline for Relikely Hand Gesture Classification of FMG Signals with Stochastic Variance. *Sensors* **2021**, *21*, 1504. [CrossRef]
37. Gurland, J.; Tripathi, R.C. A Simple Approximation for Unbiased Estimation of the Standard Deviation. *Am. Stat.* **1971**, *25*, 30–32. [CrossRef]
38. Moors, J.J.A. The Meaning of Kurtosis: Darlington Reexamined. *Am. Stat.* **1986**, *40*, 283–284. [CrossRef]
39. Doane, D.P.; Seward, L.E. Measuring Skewness: A Forgotten Statistic? *J. Stat. Educ.* **2011**, *19*, 1–18. [CrossRef]
40. Clarenz, U.; Rumpf, M.; Telea, A. Robust Feature Detection and Local Classification for Surfaces based on Moment Analysis. *IEEE Trans. Vis. Comput. Graph.* **2004**, *10*, 516–524. [CrossRef]
41. Hippenstiel, R.; Oliviera, P.D. Time-varying Spectral Estimation using the Instantaneous Power Spectrum (IPS). *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 1752–1759. [CrossRef]
42. Choi, H.; Williams, W. Improved Time-Frequency Representation of Multicomponent Signals using Exponential kernels. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 862–871. [CrossRef]
43. Loza, A.; Canagarajah, N.; Bull, D. A Simple Scheme for Enhanced Reassignment of the SPWV Representation of Noisy Signals. *IEEE Int. Symp. Image Signal Process. Anal. ISPA2003* **2003**, *2*, 630–633. [CrossRef]
44. Zhang, A.; Nowruzi, F.E.; Laganiere, R. RADDet: Range-Azimuth-Doppler based Radar Object Detection for Dynamic Road Users. *arXiv* **2021**. [CrossRef]
45. Qing, Y.; Liu, W. Hyperspectral Image Classification based on Multi-scale Residual Network with Attention Mechanism. *Remote Sens.* **2021**, *13*, 335. [CrossRef]
46. Gao, L.; Hong, D.; Yao, J.; Zhang, B.; Gamba, P.; Chanussot, J. Spectral Superresolution of Multispectral Imagery with joint Sparse and Low-rank Learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 2269–2280. [CrossRef]
47. Yin, X.; Liu, Z.; Liu, D.; Ren, X. A Novel CNN-based Bi-LSTM Parallel Model with Attention Mechanism for Human Activity Recognition with Noisy Data. *Sci. Rep.* **2022**, *12*, 7878. [CrossRef] [PubMed]

48. Wang, Z.; Bian, S.; Lei, M.; Zhao, C.; Liu, Y.; Zhao, Z. Feature Extraction and Classification of Load Dynamic Characteristics based on Lifting Wavelet Packet Transform in Power System Load Modeling. *Int. J. Electr. Power Energy Syst.* **2023**, *62*, 353–363. [CrossRef]
49. Auger, F.; Flandrin, P.; Gonçalvès, P.; Lemoine, O. Time-Frequency Toolbox For Use with MATLAB. *Tftb-Info* **2008**, *7*, 465–468. Available online: http://tftb.nongnu.org/ (accessed on 15 July 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.